ACADEMIC PRESS
Available at WWW.MATHEMATICSWEB.ORG POWERED BY SCIENCE @ DIRECT®
Journal of Multivariate Analysis 88 (2004) 118–130
Journal of Multivariate Analysis
http://www.elsevier.com/locate/jmva

An extension of the factorization theorem to the non-positive case

Pawel Kopciuszewski*

Department of Mathematics, Czestochowa University of Technology, Dabrowskiego 73, 42-200 Czestochowa, Poland

Received 14 August 2001

Abstract

This paper presents a method of determining joint distributions by known conditional distributions. A generalization of the Factorization Theorem is proposed. The generalized theorem is proved under the assumption that the support of unknown joint distribution may be divided into a countable number of sets, which all satisfy the relative weak positivity condition. This condition is defined in the paper and it generalizes the positivity condition introduced by Hammersley and Clifford. The theorem is illustrated with three examples. In the first example we determine a joint density in the case when the support of an unknown density is a continuous nonproduct set from Euclidean space $\mathscr{R}^2$. In the second example we seek the joint probability for the number of trials and the number of successes in Bernoulli's scheme. We also examine a simple example given by Kaiser and Cressie (J. Multivariate Anal. 73 (2000) 199).

AMS 2000 subject classifications: 62E10; 62H05

Keywords: Hammersley-Clifford Theorem; Positivity condition; Conditional distributions; Functionally compatible distributions

1. Introduction

The problem of determining the joint distribution by known conditional distributions has been investigated by many authors, for example by Brook [4], Spitzer [12], Hammersley and Clifford [6], Besag [2,3] and lately by Arnold and

*Fax: 0048-34-3250324.
E-mail address: pkop@imi.pcz.czest.pl.

0047-259X/03/$ - see front matter
doi:10.1016/S0047-259X(03)00055-1

Press [1], Hobert and Casella [7,8], Hobert et al. [9], Kopciuszewski [11], Kaiser and Cressie [10] and others. The theorem proved by Hammersley and Clifford [6] is the crucial theorem, which presents the formula to determine the joint distribution by conditional distributions under the positivity condition proposed by these authors. If a joint distribution is uniquely determined by conditional distributions then the only alternative way to obtain a joint distribution by means of conditional distributions is the usage of Gibbs sampling algorithm [5,7–9,11]. In the paper we introduce the relative weak positivity condition, which generalizes the positivity condition introduced by Hammersley and Clifford [6] and the MRF support condition introduced by Kaiser and Cressie [10]. Under this new condition we propose a method which makes a joint distribution dependent on conditional distributions even in the case when the support of a joint distribution is not a Cartesian product. The theorem is illustrated with three examples, where the supports of joint distributions are not Cartesian products.

2. A generalization of the factorization theorem, based on the current results

Let ν_i, $i = 1, \ldots, n$, be a σ-finite measure on the measurable space $(\mathscr{R}, \mathscr{B})$. Let $\Pi_{i|x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n}$, $i = 1, \ldots, n$, be a probability measure which depends on $n-1$ values $x_j \in \mathscr{R}$, $j = 1, \ldots, n$, $j \neq i$. Let $\pi_i(x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = \pi_i(x_i|x_j, j \neq i)$ be the density of the measure $\Pi_{i|x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n}$, with respect to the measure ν_i, $i = 1, \ldots, n$. Let $\nu = \nu_1 \times \cdots \times \nu_n$ be the product measure on the space $(\mathscr{R}^n, \mathscr{B}^n)$.

Definition 1. The densities $\pi_i(x_i|x_j, j \neq i)$, $i = 1, \ldots, n$, are said to be functionally compatible if there exists a nonnegative ν-measurable function $\pi(x_1, \ldots, x_n)$, such that for all $x \in \mathscr{R}^n$

$$\pi_i(x_i|x_j, j \neq i) \propto \pi(x_1, \ldots, x_n), \tag{2.1}$$

the function π is the joint density (proper or improper) determined by densities $\pi_1, \ldots, \pi_n$.

If $\int_{\mathscr{R}^n} \pi(x)\nu(dx) < \infty$ then the densities $\pi_1, \ldots, \pi_n$ are said to be compatible, i.e. the densities $\pi_1, \ldots, \pi_n$ are conditionals, and π is the proper joint density determined by $\pi_1, \ldots, \pi_n$.

Proportionality (2.1) concerns the variable x_i, $i = 1, \ldots, n$, only.

If the densities π_i, $i = 1, \ldots, n$, are functionally compatible then the set $S = \{x \in \mathscr{R}^n : \pi(x) > 0\}$ is the support of the joint density π. Proportionality (2.1) implies that the densities π_i, $i = 1, \ldots n$, are positive for all $x \in S$.

Hammersley and Clifford [6] introduced the positivity condition. From this condition, apart from some trivial cases, it results that the support S of density π is a Cartesian product of n sets. They gave a formula to determine the joint density $\pi(x_1, \ldots, x_n)$ by some given conditionals $\pi_i(x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$, $i = 1, \ldots, n$, under the positivity condition. Hammersley and Clifford proved the following theorem.

Theorem 1 (Hammersley–Clifford Theorem). *Under the positivity condition, the joint density* $\pi(x_1, \dots, x_n)$ *is proportional to*

$$\prod_{j=1}^{n} \frac{\pi_{r_j}(x_{r_j}|x_{r_1}, \dots, x_{r_{j-1}}, x'_{r_{j+1}}, \dots, x'_{r_n})}{\pi_{r_j}(x'_{r_j}|x_{r_1}, \dots, x_{r_{j-1}}, x'_{r_{j+1}}, \dots, x'_{r_n})}, \tag{2.2}$$

for every permutation $(r_1, \dots, r_n)$ *on* $\{1, \dots, n\}$ *and every* $x' \in S$.

This theorem is sometimes known as the Factorization Theorem [3].

Result (2.2) is satisfied, since it exists at least one point $x' \in S$ and a permutation $(r_1, \dots, r_n)$ on $\{1, \dots, n\}$ such that $\pi_{r_j}(x'_{r_j}|x_{r_1}, \dots, x_{r_{j-1}}, x'_{r_{j+1}}, \dots, x'_{r_n})$ is positive for all $x \in S$ and all $j = 1, \dots, n$. Hence, the positivity condition is too strong an assumption to determine the joint density π by the given conditionals $\pi_1, \dots, \pi_n$. For this reason we introduce a new definition of *weak positivity*.

Definition 2. Functionally compatible densities $\pi_1, \dots, \pi_n$ satisfy the weak positivity condition (in short WP) on a set $A \subset S$, if there exists a point $x' \in A$ and a permutation $(r_1, \dots, r_n)$ on $\{1, \dots, n\}$, such that for ν-almost all points $x \in A$ and all $j = 1, \dots, n$,

$$\pi_{r_j}(x'_{r_j}|x_{r_1}, \dots, x_{r_{j-1}}, x'_{r_{j+1}}, \dots, x'_{r_n}) > 0. \tag{2.3}$$

Certainly, if densities $\pi_1, \dots, \pi_n$ satisfy the positivity condition then they also satisfy the WP condition. Moreover, it should be noticed that if densities $\pi_1, \dots, \pi_n$ satisfy the WP condition on a set A, then there exists a point $x' \in A$ and a permutation $(r_1, \dots, r_n)$ on $\{1, \dots, n\}$, such that the density π satisfies result (2.2) for ν-almost all $x \in A$.

Kaiser and Cressie [10] introduced the Markov random field (MRF) support condition which means that there exists a point $x' \in A$ such that for all permutations $(r_1, \dots, r_n)$ on $\{1, \dots, n\}$ and for all points $x \in A$ condition (2.3) holds. That is why the MRF support condition given by Kaiser and Cressie is satisfied by a much smaller class of supports than the WP condition.

We illustrate the idea of the WP condition with two short examples. Let A be the circle shown in Fig. 1a and let π_1, π_2 be two functionally compatible densities, which are positive on A. Notice that there exists a point $x' \in A$ (from the diameter of the circle A, which is parallel to axis x_1) and permutation (1,2), such that for all $x = (x_1, x_2) \in A$,

$$\pi_1(x'_1|x'_2) > 0 \quad \text{and} \quad \pi_2(x'_2|x_1) > 0.$$

Hence, the densities π_1, π_2 satisfy the WP condition on the set A. Moreover, if the set A is the support of the density π, then π can be determined from result (2.2). It should be noticed that inequality (2.3) is not satisfied for the point $x' \in A$ and the

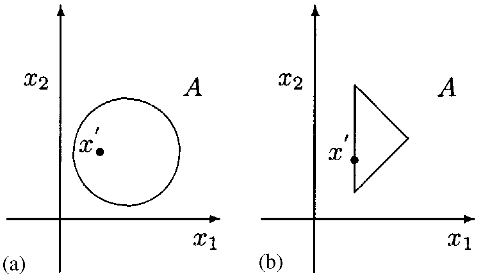

Fig. 1. a.) The densities π_1, π_2 satisfy the WP condition with the point x' and the permutation (1,2). b.) The densities π_1, π_2 satisfy the WP condition with the point x' and the permutation (2,1).

permutation (2,1), because there exist some points $x = (x_1, x_2) \in A$, for which $\pi_1(x'_1|x_2) = 0$.

Suppose now that the set A is the triangle shown in Fig. 1b. In this case there exists a point $x' \in A$ (from the side of the triangle A, which is parallel to axis x_2) and permutation (2,1), such that for two functionally compatible densities π_1 and π_2, which are positive on the set A and for all points $x = (x_1, x_2) \in A$,

$$\pi_2(x'_2|x'_1) > 0 \quad \text{and} \quad \pi_1(x'_1|x_2) > 0.$$

Therefore π_1, π_2 satisfy the WP condition on the set A. Note that the WP condition is not satisfied for the point $x' \in A$ and the permutation (1,2), because there exist some points $x = (x_1, x_2) \in A$, for which $\pi_2(x'_2|x_1) = 0$.

In [11] the relation R_A between two points $x, y \in A \subset S$ was introduced as follows.

Definition 3. A point $x \in A$ is in relation R_A to a point $y \in A$, which we write as xR_Ay, iff there exists a natural number k and a finite sequence of vectors $c^{(j)} = (0, \ldots, 0, c_{m_j}, 0, \ldots, 0), c_{m_j} \neq 0, j = 0, \ldots, k-1$, from Euclidean space $\mathscr{R}^n$, which have only one component different from zero, such that

$$\begin{aligned} &z^{(0)} = x, \\ &z^{(j+1)} = z^{(j)} + c^{(j)} \in A, \ \text{where } j = 0, \ldots, k-1, \\ &z^{(k)} = y. \end{aligned}$$

The above definition means that two points $x \in A$ and $y \in A$ can be joined by a broken line with all points within the set A and any of its intervals is parallel to one of the axes.

Let I be a subset of nonnegative integers, the least element of which is equal to zero. In particular, the set I can be a finite set, for example $I = \{0, 1, \ldots, n\}$, $0 \leqslant n < \infty$, or the set of all nonnegative integers.

Below we introduce another definition of *relative weak positivity*.

Definition 4. Functionally compatible densities $\pi_1, \dots, \pi_n$ satisfy the relative weak positivity condition (in short RWP), iff

(i) there exists a sequence of disjoint sets $S_i \subset S, i \in I$, (not necessarily with positive ν measure), such that $\pi_1, \dots, \pi_n$ satisfy the WP condition on all these sets and such that $S = \bigcup_{i \in I} S_i$; that is, for any $i \in I$ there exists a point $x^{(i)} \in S_i$ and a permutation $(r_{i,1}, \dots, r_{i,n})$ on $\{1, \dots, n\}$, such that for ν-almost all $x \in S_i$ and all $j = 1, \dots, n$

$$\pi_{r_j}(x^{(i)}_{r_{i,j}} | x_{r_{i,1}}, \dots, x_{r_{i,j-1}}, x^{(i)}_{r_{i,j+1}}, \dots, x^{(i)}_{r_{i,n}}) > 0.$$

(ii) $x^{(i)} R_S x^{(i+1)}$, for all $i, i+1 \in I$; that is, for any $i \in I$ there exists a natural number k_i and a finite sequence of vectors $c^{(i,j)} = (0, \dots, 0, c_{m_{i,j}}, 0, \dots, 0)$, $c_{m_{i,j}} \neq 0$, $j = 0, \dots, k_i - 1$, from Euclidean space $\mathscr{R}^n$, satisfying

$$z^{(i,0)} = x^{(i)},$$

$$z^{(i,j+1)} = z^{(i,j)} + c^{(i,j)} \in S, \ \text{where } j = 0, \dots, k_i - 1,$$

$$z^{(i,k_i)} = x^{(i+1)}.$$

Below we present some remarks related to this definition.

We can notice that $z^{(i,j)}$ and $z^{(i,j+1)}$ differ on their $m_{i,j}$ component, for all $i \in I$, $j = 0, \dots, k_i - 1$. From result (2.2) it follows that

$$\pi(z^{(i,j+1)}) = C_{i,j} \pi(z^{(i,j)}), \tag{2.4}$$

where

$$C_{i,j} = \frac{\pi_{m_{i,j}}(z^{(i,j+1)}_{m_{i,j}} | z^{(i,j)}_1, \dots, z^{(i,j)}_{m_{i,j}-1}, z^{(i,j)}_{m_{i,j}+1}, \dots, z^{(i,j)}_n)}{\pi_{m_{i,j}}(z^{(i,j)}_{m_{i,j}} | z^{(i,j)}_1, \dots, z^{(i,j)}_{m_{i,j}-1}, z^{(i,j)}_{m_{i,j}+1}, \dots, z^{(i,j)}_n)}.$$

In consequence, we conclude that for all $i, i+1 \in I$,

$$\pi(x^{(i+1)}) = D_i \pi(x^{(i)}), \tag{2.5}$$

where

$$D_i = \prod_{j=0}^{k_i - 1} C_{i,j}.$$

Finally, for all $i \in I$ the density π satisfies

$$\pi(x^{(i)}) = \prod_{j=0}^{i-1} D_j \pi(x^{(0)}), \tag{2.6}$$

where the notation $\prod_{j=0}^{-1} D_j = 1$ is used in the paper, for simplicity.

It is very important for our results that the above condition makes all densities $\pi(x^{(i)}), i \in I$, dependent on the density $\pi(x_0)$ in only one point $x_0 \in S$.

Now we can propose a theorem which is a generalization of the Hammersley–Clifford Theorem (known as the Factorization Theorem).

Theorem 2. *If functionally compatible densities* $\pi_1, \dots, \pi_n$ *satisfy the RWP condition, then the joint density* $\pi(x_1, \dots, x_n)$ *is proportional to*

$$\sum_{i \in I} \left\{ \prod_{j=1}^{n} \frac{\pi_{r_{i,j}}(x_{r_{i,j}} | x_{r_{i,1}}, \dots, x_{r_{i,j-1}}, x^{(i)}_{r_{i,j+1}}, \dots, x^{(i)}_{r_{i,n}})}{\pi_{r_{i,j}}(x^{(i)}_{r_{i,j}} | x_{r_{i,1}}, \dots, x_{r_{i,j-1}}, x^{(i)}_{r_{i,j+1}}, \dots, x^{(i)}_{r_{i,n}})} \prod_{j=0}^{i-1} D_j 1_{S_i}(x) \right\}. \tag{2.7}$$

Proof. From item (i) of the RWP definition it follows that there exists a disjoint countable covering $\{S_i\}_{i \in I}$ of the set S such that for any $i \in I$ there exists a permutation $(r_{i,1}, \dots, r_{i,n})$ on $\{1, \dots, n\}$ and a point $x^{(i)} \in S_i$, such that for ν-almost all $x \in S_i$ and all $j = 1, \dots, n$,

$$\pi_{r_{i,j}}(x^{(i)}_{r_{i,j}} | x_{r_{i,1}}, \dots, x_{r_{i,j-1}}, x^{(i)}_{r_{i,j+1}}, \dots, x^{(i)}_{r_{i,n}}) > 0.$$

From result (2.2) it follows that for ν-almost all $x \in S_i$ the density $\pi(x)$ is equal to

$$\prod_{j=1}^{n} \frac{\pi_{r_{i,j}}(x_{r_{i,j}} | x_{r_{i,1}}, \dots, x_{r_{i,j-1}}, x^{(i)}_{r_{i,j+1}}, \dots, x^{(i)}_{r_{i,n}})}{\pi_{r_{i,j}}(x^{(i)}_{r_{i,j}} | x_{r_{i,1}}, \dots, x_{r_{i,j-1}}, x^{(i)}_{r_{i,j+1}}, \dots, x^{(i)}_{r_{i,n}})} \pi(x^{(i)}). \tag{2.8}$$

From (2.6), which is the conclusion from item (ii) of the RWP definition and from (2.8) we have that for all $i \in I$ and ν-almost all $x \in S_i$ the density $\pi(x)$ is equal to

$$\prod_{j=1}^{n} \frac{\pi_{r_{i,j}}(x_{r_{i,j}} | x_{r_{i,1}}, \dots, x_{r_{i,j-1}}, x^{(i)}_{r_{i,j+1}}, \dots, x^{(i)}_{r_{i,n}})}{\pi_{r_{i,j}}(x^{(i)}_{r_{i,j}} | x_{r_{i,1}}, \dots, x_{r_{i,j-1}}, x^{(i)}_{r_{i,j+1}}, \dots, x^{(i)}_{r_{i,n}})} \prod_{j=0}^{i-1} D_j \pi(x^{(0)}).$$

The sets $S_i, i \in I$, are disjoint covering of the set S. Hence it is easily seen that (2.7) is proved for ν-almost all $x \in S$.

Now we define the neighborhood of any site $i \in \{1, \dots, n\}$ and formulate the straightforward corollary to Theorem 2. Site $j \neq i$ is said to be a neighbor of site i if and only if the density $\pi_i(x_i | x_1, \dots x_{i-1}, x_{i+1}, \dots, x_n)$ depends functionally upon the variable x_j. Notice that the density $\pi_{i_1, \dots, i_l}$ of the distribution of any random vector $(X_{i_1}, \dots, X_{i_l})^T, i_1, \dots, i_l \in \{1, \dots, n\}, l \geqslant 1$, given all other site values can be determined by conditionals $\pi_{i_1}, \dots, \pi_{i_l}$, using (2.7) if these conditionals satisfy the RWP condition (on the support of the density $\pi_{i_1, \dots, i_l}$). Then the corollary from Theorem 2 is that the distribution of any random vector $(X_{i_1}, \dots, X_{i_l})^T$ given all other site values depends only upon the realizations $x_{i_1}, \dots, x_{i_l}$ of the variables $X_{i_1}, \dots, X_{i_l}$ and the values at neighboring sites $i_1, \dots, i_l$.

3. More convincing examples

In the next three examples we find the joint distributions with nonproduct supports by means of Theorem 2.

Notice that two densities $\pi_1(x_1|x_2)$ and $\pi_2(x_2|x_1)$ are functionally compatible iff there exist two functions f_1 and f_2 positive on S, such that for all $(x_1, x_2) \in S$,

$$\frac{\pi_1(x_1|x_2)}{\pi_2(x_2|x_1)} = f_1(x_1) f_2(x_2).$$

3.1. Determining the continuous joint density with nonproductive support

Suppose that

$$\pi_2(x_2|x_1) \propto \exp\left\{-\frac{(x_2 + x_1 - 0.5)^2}{2}\right\},$$

$$\pi_1(x_1|x_2) \propto exp\{-x_1 x_2\},$$

where $x_1 \geqslant 0, 0 \leqslant x_2 - x_1 \leqslant 1$.

From the remark at the beginning of this section it is easily seen that the densities π_1 and π_2 are functionally compatible. Notice that π_2 is proportional to the normal density $N(-x_1 + 0.5; 1)$ on the interval $[x_1, x_1 + 1]$ and π_1 is proportional to the exponential density with parameter x_2 on the interval $[x_2 - 1, x_2]$.

Then the support S of the unknown joint density π is not a product set. The support S is the set included between two straight lines $x_2 = x_1$ and $x_2 = x_1 + 1$; that is, $S = \{x = (x_1, x_2) \in \mathscr{R}^2 \colon x_1 \geqslant 0, 0 \leqslant x_2 - x_1 \leqslant 1\}$ (Fig. 2).

Notice that there exists a sequence of disjoint sets $S_i, i = 0, 1, \ldots$ such that

(i) $S_i = \{(x_1, x_2) \in S \colon x_1 \in [i, i+1), x_2 \in [x_1, x_1 + 1]\}$ and $S = \bigcup_{i=0}^{\infty} S_i$,

(ii) there exists a sequence of points $x^{(i)} = (i, i+1) \in S_i, i = 0, 1, \ldots$, such that for all $x \in S_i$,

$$\pi_1(x_1^{(i)}|x_2^{(i)}) > 0 \quad \text{and} \quad \pi_2(x_2^{(i)}|x_1) > 0;$$

that is, the densities π_1 and π_2 fulfil the WP condition on the sets $S_i, i = 0, 1, \ldots$ (Fig. 2).

(iii) $x^{(i)} R_S x^{(i+1)}$, for all $i \geqslant 0$ (see the comment under Definition 3).

From Theorem 2 it results that the unknown joint density π is proportional to

$$\sum_{i=0}^{\infty} \left\{ \frac{\pi_2(x_2|x_1)\pi_1(x_1|x_2^{(i)})}{\pi_2(x_2^{(i)}|x_1)\pi_1(x_1^{(i)}|x_2^{(i)})} \prod_{j=0}^{i-1} D_j 1_{S_i}(x) \right\}. \tag{3.1}$$

Firstly, we find the product $\prod_{j=0}^{i-1} D_j, i = 1, 2, \ldots$. From (2.5) we have

$$\pi(x^{(i+1)}) = D_i \pi(x^{(i)}), \ i = 0, 1, \ldots,$$

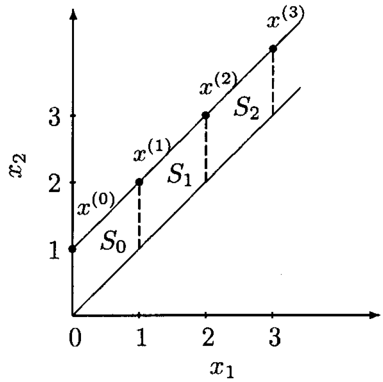

Fig. 2. The densities π_1, π_2 satisfy the WP condition on the set S_0 with the point $x^{(0)}$; on the set S_1 with the point $x^{(1)}$ and on the set S_2 with the point $x^{(2)}$, etc.

where

$$D_i = \frac{\pi_2(x_2^{(i+1)}|x_1^{(i+1)})\pi_1(x_1^{(i+1)}|x_2^{(i)})}{\pi_2(x_2^{(i)}|x_1^{(i+1)})\pi_1(x_1^{(i)}|x_2^{(i)})}.$$

The components of the points $x^{(i)}$ and $x^{(i+1)}$ satisfy $x_1^{(i+1)} = x_2^{(i)}$ and $x_2^{(i+1)} = x_2^{(i)} + 1, i = 0, 1, \ldots$. Hence, $\prod_{j=0}^{i-1} D_j$ is equal to

$$\frac{\pi_2(2|1)\pi_1(1|1) \cdot \ldots \cdot \pi_2(i+1|i)\pi_1(i|i)}{\pi_2(1|1)\pi_1(0|1) \cdot \ldots \cdot \pi_2(i|i)\pi_1(i-1|i)}.$$

It implies that $\prod_{j=0}^{i-1} D_j$ is equal to

$$\frac{\exp\left\{-\frac{(2+0.5)^2}{2}\right\}\exp\{-1^2\} \cdot \ldots \cdot \exp\left\{-\frac{(2i+0.5)^2}{2}\right\}\exp\{-i^2\}}{\exp\left\{-\frac{(2-0.5)^2}{2}\right\}\exp\{-0 \cdot 1\} \cdot \ldots \cdot \exp\left\{-\frac{(2i-0.5)^2}{2}\right\}\exp\{-(i-1)i\}}.$$

Finally, $\prod_{j=0}^{i-1} D_j, i = 1, 2, \ldots,$ is equal to

$$\prod_{j=1}^{i} \frac{\exp\left\{-\frac{(2j+0.5)^2}{2}\right\}\exp\{-j^2\}}{\exp\left\{-\frac{(2i-0.5)^2}{2}\right\}\exp\{-(j-1)j\}} = \prod_{j=1}^{i} \exp\{-3j\} = \exp\left\{-\frac{3}{2}(i+1)i\right\}. \tag{3.2}$$

From (3.1) and (3.2) we conclude that the density π is proportional to

$$\sum_{i=0}^{\infty}\left\{\frac{\exp[-\frac{1}{2}(x_2+x_1-0.5)^2]\exp[-x_1(i+1)]}{\exp[-\frac{1}{2}(i+1+x_1-0.5)^2]\exp[-i(i+1)]}\exp[-\frac{3}{2}(i+1)i]1_{S_i}(x)\right\}.$$

Hence the joint density π is proportional to

$$\exp\left\{-\frac{1}{2}x_2(x_2+2x_1-1)\right\}.$$

It is easily seen that the joint density π is a continuous function in all points $(i, x_2)\in S, i = 0, 1, \ldots,$ on the edge of the set S_i. Therefore, it is a continuous function in all $x\in S$.

3.2. *Determining the joint distribution of the number of trials and the number of succeses in Bernoulli's scheme*

Suppose that we observe $n\geqslant 1$ independent random variables X_i, $i = 1, \ldots, n$, with identical zero-one distribution, where n is the realization of a random variable N. The distribution of $X = \sum_{i=1}^{n} X_i$ is then the Binomial distribution $B(n, p)$ with parameters n and $p = P(X_i = 1) > 0$, $i = 1, \ldots, n$. It follows that the conditional distribution $P_1(X = x|N = n) = P_1(x|n)$ is the Binomial distribution. Moreover, assume that the conditional distribution $P_2(N = n|X = x) = P_2(n|x)$ is equal to $\frac{\lambda_x^{n-x}}{(n-x)!}\exp\{-\lambda_x\}, \lambda_x > 0,\ n = x, x+1, \ldots$. It is the Poisson distribution for the variable $N - x$ with parameter $\lambda_x > 0$.

Note that the conditionals P_1 and P_2 satisfy

$$\frac{P_1(x|n)}{P_2(n|x)} = n!(x!)^{-1}p^x(1-p)^{n-x}\lambda_x^{-n}\lambda_x^x\exp\{\lambda_x\}.$$

Hence, $P_1(x|n)$ and $P_2(n|x)$ are functionally compatible iff there exist two functions g_1, g_2, such that $\lambda_x^{-n} = g_1(x)g_2(n)$ for all $n = 1, 2, \ldots$ and $x = 0, 1, \ldots, n$. It is easily seen that λ_x is a constant, which is independent on the value x.

The Poisson distribution $P_2(n|x)$ is the only distribution from the well-known class of distributions which is functionally compatible with Binomial distribution $P_1(x|n)$ (if λ_x is a constant with respect to x).

The main aim of this example is to obtain the joint distribution $P(X = x, N = n) = P(x, n)$. The support S (Fig. 3) of the joint distribution P satisfies the following condition:

$$S = \{(x, n)\in\mathscr{R}^2 : n = 1, 2, \ldots; x = 0, 1, \ldots, n\}.$$

Hence, the conditional distributions P_1 and P_2 do not fulfil the positivity condition. They satisfy the WP condition on the whole set S with the point $x^{(0)} = (x, n) = (0, 1)$ and with the permutation (2,1). According to the WP condition we have for all $n = 1, 2, \ldots$

$$P_1(0|n) > 0 \quad \text{and} \quad P_2(1|0) > 0.$$

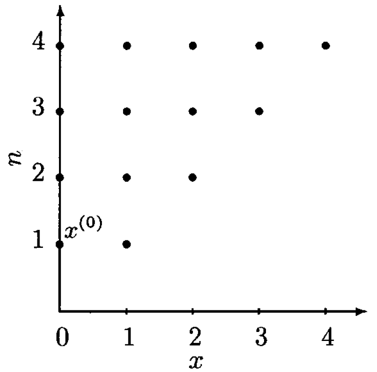

Fig. 3. The support S of the random vector (X, N).

Notice that there does not exist a point $(x', n') \in S$, for which the distributions P_1 and P_2 fulfil the WP condition on the set S with the permutation (1,2), because $P_2(n'|x) = 0$ for all $n' = 1, 2, \ldots$ and for all $x = n' + 1, n' + 2, \ldots$.

Then, the joint distribution $P(x, n)$ is proportional to

$$\frac{P_1(x|n)P_2(n|0)}{P_1(0|n)P_2(1|0)}.$$

Hence, $P(x, n)$ satisfies

$$P(x, n) \propto \frac{1}{x!(n-x)!}\left(\frac{p}{1-p}\right)^x \lambda^n.$$

Finally, $P(x, n), n = 1, 2, \ldots, x = 0, 1, \ldots, n,$ is the proper distribution, which is equal to

$$\left(\exp\left\{\frac{\lambda}{1-p}\right\} - 1\right)^{-1} \frac{1}{x!(n-x)!}\left(\frac{p}{1-p}\right)^x \lambda^n.$$

That is why P_1 and P_2 are compatible.

3.3. Two finite discrete models

The first example is the one presented by Kaiser and Cressie [10]. The authors considered the conditionals which did not satisfy the MRF support condition and that is why the joint distribution could not be determined by their theorem.

Suppose that we have two random variables $X_1 \in \{0, 1\}$ and $X_2 \in \{1, 2, 3\}$ and consider the conditional specifications:

$$P_1(0|1) = 1, P_1(0|2) = 0.43, P_2(1|0) = 0.7,$$

$$P_2(2|0) = 0.3, P_2(2|1) = 0.4, P_2(3|1) = 0.6,$$

where $P_i(x_i|x_j) = P_i(X_i = x_i|X_j = x_j), i, j = 1, 2.$

For these conditionals $S=\{(0,1),(0,2),(1,2),(1,3)\}$ and it is easily seen that the conditionals satisfy the RWP condition on the set S (but not the MRF support condition).

(i) There exist two disjoint sets $S_1=\{(0,1),(0,2)\}$ and $S_2=\{(1,2),(1,3)\}$, such that the conditionals P_1 and P_2 satisfy the WP condition (even the positivity condition) on the sets S_1 and S_2. That is, there exist two points $x^{(0)}=(0,2)\in S_1$ and $x^{(1)}=(1,2)\in S_2$ such that

$$P_2(X_2=x_2|X_1=0)>0 \quad \text{for} \quad (x_1,x_2)\in S_1 \text{ and}$$

$$P_2(X_2=x_2|X_1=1)>0 \quad \text{for} \quad (x_1,x_2)\in S_2.$$

(ii) $x^{(0)}R_S x^{(1)}$ because the points $x^{(0)}$ and $x^{(1)}$ differ on their first component, only.

From formula (2.7) we have

$$P(X_1=0,X_2=1)=0.35, P(X_1=0,X_2=2)=0.15,$$

$$P(X_1=1,X_2=2)=0.2, P(X_1=1,X_2=3)=0.3.$$

Now suppose that we have four random variables $X_1,X_2,X_3,X_4\in\{0,1\}$ and consider the conditional specifications:

$$P_1(0|0,0,0)=a>0, P_1(1|0,0,0)=1-a>0, P_2(0|1,0,0)=b>0,$$

$$P_2(1|1,0,0)=1-b>0, P_3(0|1,1,0)=c>0, P_3(1|1,1,0)=1-c>0,$$

$$P_4(0|1,1,1)=d>0, P_4(1|1,1,1)=1-d>0,$$

where $P_i(X_i=x_i|X_j=x_j,X_k=x_k,X_l=x_l)=P_i(x_i|x_j,x_k,x_l)$ and $i,j,k,l\in\{1,2,3,4\}$ are different sites.

For these conditionals $S=\{(0,0,0,0),(1,0,0,0),(1,1,0,0),(1,1,1,0),(1,1,1,1)\}$. Certainly, the conditionals satisfy the RWP condition, because

(i) There exist two disjoint sets $S_1=\{(0,0,0,0),(1,0,0,0),(1,1,0,0)\}$ and $S_2=\{(1,1,1,0),(1,1,1,1)\}$, such that the conditionals satisfy the WP condition on the sets S_1 and S_2. That is, there exist two points $x^{(0)}=(1,0,0,0)\in S_1$ and $x^{(1)}=(1,1,1,0)\in S_2$ such that

$$P_2(0|x_1,x_3,x_4)>0, P_1(1|0,x_3,x_4)>0 \text{ for } (x_1,x_2,x_3,x_4)\in S_1 \text{ and}$$

$$P_3(1|x_1,x_2,x_4)>0, P_4(1|x_1,x_2,1)>0 \text{ for } (x_1,x_2,x_3,x_4)\in S_2.$$

(ii) $x^{(0)}R_S x^{(1)}$ (see the comment under Definition 3).

From formula (2.5) or (2.6) we have

$$P(x^{(1)})=\frac{P_3(1|1,1,0)P_2(1|1,0,0)}{P_3(0|1,1,0)P_2(0|1,0,0)}P(x^{(0)}).$$

From formula (2.7) we have

$$P(0,0,0,0) = \frac{P_1(0|0,0,0)}{P_1(1|0,0,0)} P(x^{(0)}), \;\; P(1,1,0,0) = \frac{P_2(1|1,0,0)}{P_2(0|1,0,0)} P(x^{(0)}),$$

$$P(1,1,1,1) = \frac{P_4(1|1,1,1)}{P_4(0|1,1,1)} \frac{P_3(1|1,1,0)P_2(1|1,0,0)}{P_3(0|1,1,0)P_2(0|1,0,0)} P(x^{(0)}).$$

Hence,

$$P(x^{(1)}) = P(1,1,1,0) = \frac{(1-c)(1-b)}{cb} P(x^{(0)}),$$

$$P(0,0,0,0) = \frac{a}{1-a} P(x^{(0)}), P(1,1,0,0) = \frac{1-b}{b} P(x^{(0)}),$$

$$P(1,1,1,1) = \frac{(1-d)(1-c)(1-b)}{dcb} P(x^{(0)}).$$

For example $a = 0.5$, $b = 0.4$, $c = 0.3$, $d = 0.2$. Then we have

$$P(x^{(0)}) = P(1,0,0,0) = \tfrac{1}{21}, P(0,0,0,0) = \tfrac{1}{21}, P(1,1,0,0) = \tfrac{1}{14},$$

$$P(1,1,1,0) = \tfrac{1}{6}, P(1,1,1,1) = \tfrac{2}{3}.$$

Generally, it seems that difficulties with the construction of joint distributions are connected only with finding a covering S_i, $i = 0, 1, \ldots,$ of the support S such that the RWP condition is satisfied, not with the dimensionality of the support.

Acknowledgments

The author is grateful to the editors and referees for many constructive comments that lead to substantial improvements.

References

[1] B.S. Arnold, J. Press, Compatible conditional distributions, J. Amer. Statist. Assoc. 84 (1989) 152–156.

[2] J.E. Besag, Nearest-neighbour systems and the auto-logistic model for binary data, J. Roy. Statist. Soc. B 34 (1972) 75–83.

[3] J.E. Besag, Spatial interaction and the statistical analysis of lattice systems (with discussion), J. Roy. Statist. Soc. B 36 (1974) 192–236.

[4] D. Brook, On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems, Biometrika 51 (1964) 481–483.

[5] K.S. Chan, Asymptotic behaviour of the Gibbs sampler, J. Amer. Statist. Assoc. 88 (1993) 320–326.

[6] J.M. Hammersley, P. Clifford, Markov fields on finite graphs and lattices, 1971 (unpublished).

[7] J.P. Hobert, G. Casella, The effect of improper priors on Gibbs sampling in hierarchical linear mixed models, J. Amer. Statist. Assoc. 91 (1996) 1461–1473.

[8] J.P. Hobert, G. Casella, Functional compatibility, Markov chains and Gibbs sampling with improper priors, J. Comput. Graph. Statist. 7 (1998) 44–60.

[9] J.P. Hobert, C.P. Robert, C. Goutis, Connectedness conditions for the convergence of the Gibbs sampler, Statist. Probab. Lett. 33 (1997) 235–240.

[10] M.S. Kaiser, N. Cressie, The construction of multivariate distributions from Markov random fields, J. Multivariate Anal. 73 (2000) 199–220.
[11] P. Kopciuszewski, The class of joint densities uniquely determined by conditionals used in Gibbs sampling, Demonstratio Math. 32 (1999) 847–860.
[12] F. Spitzer, Markov random fields and Gibbs ensembles, Amer. Math. Monthly 78 (1971) 142–154.

Further reading

A.E. Gelfand, A.F.M. Smith, Sampling-based approaches to calculating marginal densities, J. Amer. Statist. Assoc. 85 (1990) 398–409.
A. Gelman, T.P. Speed, Characterizing a joint probability distribution by conditionals, J. Roy. Statist. Soc. B 55 (1993) 185–188.